

AD-A078 583

BROWN UNIV PROVIDENCE R I LEFSCHETZ CENTER FOR DYNAM--ETC F/G 12/1
ON THE WEAK CONVERGENCE OF A SEQUENCE OF GENERAL STOCHASTIC DIF--ETC(U)
OCT 79 H J KUSHNER , H HUANG
N00014-76-C-0279

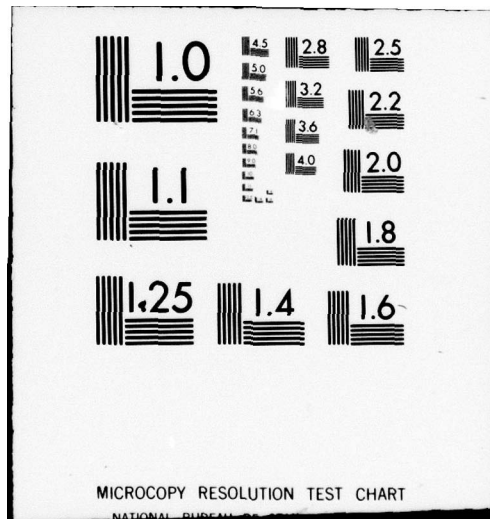
UNCLASSIFIED

LCDS-TR-79-6

NL

| OF |
ADA
078583





Lefschetz Center for Dynamical Systems



70 11

(14) LCDS-TR-79-6

6

ON THE WEAK CONVERGENCE OF A SEQUENCE OF GENERAL STOCHASTIC
DIFFERENCE EQUATIONS TO A DIFFUSION,

10

H. J. / Kushner⁺
Hai / Huang⁺⁺

12

36

ABSTRACT

11

Oct 79

A convenient and useful method for showing weak convergence, to a diffusion, of the interpolated solutions of a (not necessarily Markovian) sequence of stochastic difference equations is developed. The technique involves the use of averaging methods to show that the weak limit satisfies the martingale problem of Strook and Varadhan which is associated with the diffusion. A truncation method is developed so that it is only necessary to work with the parts of the process before first escape from an arbitrary but bounded domain. The assumptions cover a wide variety of applications in systems theory, mathematical biology and elsewhere but the method of proof is adaptable to other special cases where our particular assumptions might not hold. Two applications are given in order to illustrate the relative ease of use of the method. The driving noise process in the difference equations can depend on the solution process of the difference equation, and one application where this is useful is given (a rate of convergence problem for simple stochastic approximations with sequentially averaged observations).

*Brown University, Division of Applied Mathematics and Engineering. Lefschetz Center for Dynamical Systems. Research was supported by the Air Force Office of Scientific Research, AFOSR 77-3063, Office of Naval Research N0014-76-C-0279-P0002 and the National Science Foundation under NSF-Eng. 77-12946.

**Brown University, Division of Applied Mathematics, Lefschetz Center for Dynamical Systems. Research was supported by the National Science Foundation under NSF-Eng. 77-12946 and in part by the Office of Naval Research under N0014-76-C-0279-P0002. ✓ AFOSR-77-3063

OCTOBER, 1979

DISTRIBUTION STATEMENT A

Approved for public release;

401834

gum

| | |
|--------------------------|-------------------------------------|
| Accession For | |
| NTIS GRA&I | <input checked="" type="checkbox"/> |
| DDC TAB | <input type="checkbox"/> |
| Unannounced | <input type="checkbox"/> |
| Justification | |
| By <u>Rev H. on file</u> | |
| Distribution/ | |
| Availability Codes | |
| Dist. | Avail and/or special |
| A | |

1. Introduction

The paper develops a general method for proving weak convergence to a diffusion process of the sequence of appropriately scaled and interpolated solutions to the (not necessarily Markovian) equation

$$(1.1) \quad X_{n+1}^\epsilon = X_n^\epsilon + \epsilon h_\epsilon(X_n^\epsilon, \xi_n^\epsilon) + \sqrt{\epsilon} g_\epsilon(X_n^\epsilon, \xi_n^\epsilon) + o(\epsilon), \quad X_n^\epsilon \in \mathbb{R}^r, \\ X_0^\epsilon \text{ given.}$$

The $\{\xi_n^\epsilon\}$ are a sequence of random variables whose distributions might depend on the $\{X_n^\epsilon\}$. The method is convenient to use and has wide applicability. In order to illustrate the use of the method, applications to the rate of convergence for a general form of stochastic approximation and to a problem of Guess and Gillespie [1] are treated. For the latter problem, the treatment in [1] required an explicit construction of the solution - essentially limiting the treatment to the scalar case. There is no such restriction here.

Define $X^\epsilon(\cdot)$ by $X^\epsilon(t) = X_n^\epsilon$ on $[n\epsilon, (n+1)\epsilon)$. The basic idea is to prove that $\{X^\epsilon(\cdot)\}$ converges weakly to the solution of the martingale problem [2] connected with the diffusion process. In recent years many nice results for dealing with weak convergence of a sequence of non-Markov continuous parameter processes to a Markov process have been developed [3]-[5], but the discrete parameter case is not in such good shape.

The basic background theorems are in [6] where some "continuous parameter" applications are given. That reference emphasizes the continuous parameter case. But the method is often easier to use and can handle many types of interesting problems in the discrete parameter

case, and here we show how to effectively use it for a broad class of problems. The method of proof is interesting in itself and can be adapted to special cases when our assumptions do not hold. For each $\epsilon > 0$, let \mathcal{G}_t^ϵ be an increasing sequence of σ -algebras which measures $\{X^\epsilon(s), s \leq t\}$ and E_t^ϵ the corresponding conditional expectation operator. Write $E_{n\epsilon}^\epsilon = E_n^\epsilon$. Let $\bar{\mathcal{L}}^\epsilon$ denote the measurable functions $f(\cdot)$ (of (ω, t)) which are constant on the $[n\epsilon, n\epsilon + \epsilon)$ intervals and \mathcal{F}_t^ϵ measurable at time t , and satisfy $\sup_{t \geq 0} E|f(t)| < \infty$. Let $f^\epsilon(\cdot) \in \bar{\mathcal{L}}^\epsilon$. If $\sup_{\substack{t \geq 0 \\ \epsilon > 0}} E|f^\epsilon(t)| < \infty$ and $\lim_{\epsilon \rightarrow 0} E|f^\epsilon(t)| = 0$, for each t , we say

(following the terminology in [3], [6]) that $p\text{-}\lim_{\epsilon \rightarrow 0} f^\epsilon(\cdot) = 0$. Define \hat{A}^ϵ on $\bar{\mathcal{L}}^\epsilon$ by

$$\hat{A}^\epsilon f(t) = [E_t^\epsilon f(t+\epsilon) - f(t)]/\epsilon.$$

\hat{A}^ϵ is an approximation "in some sense" of the weak infinitesimal operator of the limit process.

Reference [3] contains a very interesting method for the continuous parameter case, with some remarks on how it might be used for the discrete parameter case. The method here seems easier to use and it is easier to construct the perturbation $\{f^\epsilon(\cdot)\}$ without method. Some of our results were strongly motivated by the techniques in [3].

Section 2 contains some assumptions on the limit process, and a sequence of truncated processes is introduced in Section 3. The use of these truncated processes will facilitate the tightness proof, and allows us to work only with $X^\epsilon(\cdot)$ and the limit until the first escape time from an arbitrary but bounded region. The general background limit theorem and the tightness theorem from [6] are stated in Section 4 in the form which will be most useful to us. In Section 5,

specific assumptions for (1.1) are given when $\{\xi_n^\epsilon\}$ does not depend on $\{X_n^\epsilon\}$, and the theorems of Section 4 are applied to (1.1) in Section 6. Sections 7 and 8 illustrate the general method with two applications. The modifications when $\{\xi_n^\epsilon\}$ is $\{X_n^\epsilon\}$ dependent are discussed in Section 9, and an application to the rate of convergence problem for a stochastic approximation with "averaged" observations appears in Section 10. From a notational point of view, it is much simpler to treat the case of non-state-dependent noise first.

2. Assumptions on the Limit Process

Some assumptions on (what will be) the limit process is required.

Let $\hat{\mathcal{L}}$ denote the real valued functions on $R^+ \times R^r$ which are zero at ∞ , $\hat{\mathcal{L}}_0$ the subset with compact support, and $\hat{\mathcal{L}}_0^{\alpha, \beta}$ the subset whose mixed partial (t, x) derivatives up to orders (α, β) are continuous. Let $A = \sum_i b_i(x, t) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i, j} a_{ij}(x, t) \frac{\partial^2}{\partial x_i \partial x_j}$ denote a diffusion operator with continuous coefficients.

Next, an existence and uniqueness condition is needed. $D^r[0, \infty)$ denotes the usual space [7] of R^r valued functions which have left hand limits and are right continuous and with the Skorokhod topology. Let $x(\cdot)$ denote the generic element of $D^r[0, \infty)$. For each $x \in R^r$, we assume that there is measure P_x on $D^r[0, \infty)$ such that $P_x\{x(0) = x\} = 1$ and

$$(2.1) \quad P_x\{\sup_{t \leq T} |x(t)| < \infty\} = 1, \text{ each } T < \infty,$$

and which is the unique solution to the martingale problem of Strook and Varadhan; namely, for each $f(\cdot, \cdot) \in \hat{\mathcal{L}}_0^{1, 2}$ and $x \in R^r$, the $M_f(\cdot)$ below is a P_x martingale:

$$(2.2) \quad M_f(t) = f(x(t), t) - f(x(0), 0) - \int_0^t \left(\frac{\partial}{\partial s} + A \right) f(x(s), s) ds.$$

We work on $D^{\mathbb{R}}[0, \infty)$ rather than on $C^{\mathbb{R}}[0, \infty)$ because it is easier to prove tightness on the former space. The measure P_x is concentrated on the subset of $D^{\mathbb{R}}[0, \infty)$ of continuous functions, in any case. When the measure is given, the corresponding process solution to the martingale problem will be written as $X(\cdot)$ in order to distinguish it from the generic element $x(\cdot)$. We still have existence and uniqueness if the initial value x is replaced by a random variable $X(0)$. Below $X(0)$ will be the weak limit of $\{X_0^{\epsilon}\}$.

3. Truncated Processes

The idea of the proof in [6] is to first prove tightness and then to show that all weak limits solve the same martingale problem whose process solution $X(\cdot)$ is unique (in the sense of measure, of course). To facilitate the proof of tightness, it is convenient to bound the random functions $X^{\epsilon}(\cdot)$ and $X(\cdot)$ by altering them after they first leave the sphere $S_N = \{x: |x| \leq N\}$ and stopping them after first exit from S_{N+1} . It is then proved that for each N the sequence of $X^{\epsilon}(\cdot)$ before first exit from S_N converges weakly to the part of the diffusion $X(\cdot)$ before first exit from S_N . Finally, the uniqueness and (2.1) (i.e., infinite escape time for the diffusion) is used to get that the unaltered $X^{\epsilon}(\cdot)$ converge weakly to $X(\cdot)$ as desired. Let $q_N(\cdot)$ denote a continuous function which takes the value unity on S_N , zero on $\mathbb{R}^{\mathbb{R}} - S_{N+1}$, has values in $[0, 1]$ and first and second

derivatives uniformly bounded in x, N . It is convenient to write (1.1) in the slightly expanded form

$$(3.1) \quad X_{n+1}^\epsilon = X_n^\epsilon + \epsilon \bar{h}_\epsilon(X_n^\epsilon) + \sqrt{\epsilon} g_\epsilon(X_n^\epsilon, \xi_n^\epsilon) + \epsilon \tilde{h}_\epsilon(X_n^\epsilon, \xi_n^\epsilon) + o(\epsilon) k_\epsilon(X_n^\epsilon, \xi_n^\epsilon),$$

where conditions on the functions will be given below.

Let the subscript N denote multiplication by $q_N(\cdot)$; i.e., $\bar{h}_{\epsilon, N}(\cdot) = \bar{h}_\epsilon(\cdot) q_N(\cdot)$, etc. For each N , we define the truncated or altered process $X_n^{\epsilon, N}$ by

$$(3.2) \quad \begin{aligned} X_{n+1}^{\epsilon, N} = & X_n^{\epsilon, N} + \epsilon \bar{h}_{\epsilon, N}(X_n^{\epsilon, N}) + \sqrt{\epsilon} g_{\epsilon, N}(X_n^{\epsilon, N}, \xi_n^\epsilon) \\ & + \epsilon \tilde{h}_{\epsilon, N}(X_n^{\epsilon, N}, \xi_n^\epsilon) + o(\epsilon) k_{\epsilon, N}(X_n^{\epsilon, N}, \xi_n^\epsilon), \end{aligned}$$

The sequence $\{X_n^{\epsilon, N}\}$ equals $\{X_n^\epsilon\}$ until at least the first time that the latter exits from S_N .

Let A^N denote a diffusion operator of the form of A in Section 3 and whose coefficients $a_N(\cdot, \cdot)$ and $b_N(\cdot, \cdot)$ are continuous and equal ^{to} $a(\cdot, \cdot)$ and $b(\cdot, \cdot)$, resp., in S_N . Suppose that a process $X^N(\cdot)$ solves (not necessarily uniquely) the martingale problem corresponding to operator A^N and (perhaps random) initial condition $X^N(0)$. If $X^N(0) \rightarrow X(0)$ weakly, then we call $X^N(\cdot)$ an N -truncation of $X(\cdot)$, the unique solution of the martingale problem for initial condition $X(0)$ and with operator A . The terms $\hat{A}^{\epsilon, N}$, $\mathcal{A}_t^{\epsilon, N}$, $E_t^{\epsilon, N}$, $E_n^{\epsilon, N}$ and $\mathcal{Z}^{\epsilon, N}$ are defined analogously to \hat{A}^ϵ , etc., but for the process $X^{\epsilon, N}(\cdot)$ instead of the process $X^\epsilon(\cdot)$, where $X^{\epsilon, N}(\cdot)$ is the piecewise constant (on the $[n\epsilon, (n+1)\epsilon]$ intervals) interpolation of $\{X_n^{\epsilon, N}\}$; in particular $X_n^{\epsilon, N} = X^{\epsilon, N}(n\epsilon)$.

4. The Limit Theorems

Theorems 1 and 2 are taken from [6], but are rephrased for the convenience of the class of problems dealt with here. In the following sections, they will be applied to (1.1) and (3.1).

In Theorems 1 and 2, $\{X_n^\epsilon, n \geq 0\}$ is an arbitrary sequence for each $\epsilon > 0$, with interpolation denoted by $X^\epsilon(\cdot)$ ($X^\epsilon(t) = X_n^\epsilon$ on $[n\epsilon, (n+1)\epsilon)$). It need not be defined by (1.1) or (3.1). The $\{X_n^{\epsilon, N}\}$ is any sequence such that $X_n^{\epsilon, N} = X_n^\epsilon$ until at least the first exit time of $\{X_n^\epsilon\}$ from S_N . Part of our basic structure (e.g., use of \hat{A}^ϵ and the perturbed $f^{\epsilon, N}(\cdot)$ below) is motivated by that of Kurtz [3]. The proofs are much different and do not require the semigroup machinery of [3] and its predecessors. An extensive development of a martingale approach to limit theorems for a class of continuous parameter Markov processes appears in [12].

Theorem 1. Assume the conditions of Sections 2 and 3 on the martingale problem with operator A , and on $a_N(\cdot, \cdot)$ and $b_N(\cdot, \cdot)$. Let $X_0^\epsilon \rightarrow X_0$ weakly as $\epsilon \rightarrow 0$. For each N and $f(\cdot, \cdot) \in \mathcal{D}$, a dense set in \mathcal{D}_0 , let there be a sequence $\{f^{\epsilon, N}(\cdot)\}$, where $f^{\epsilon, N}(\cdot) \in \mathcal{L}^{\epsilon, N}$ and such that

$$(4.1) \quad p\text{-}\lim_{\epsilon \rightarrow 0} [f^{\epsilon, N}(\cdot) - f(X^{\epsilon, N}(\cdot), \cdot)] = 0$$

$$(4.2) \quad p\text{-}\lim_{\epsilon \rightarrow 0} [\hat{A}^{\epsilon, N} f^{\epsilon, N}(\cdot) - (\frac{\partial}{\partial t} + A^N) f(X^{\epsilon, N}(\cdot), \cdot)] = 0.$$

Then, if $\{X^{\epsilon, N}(\cdot), \epsilon > 0\}$ is tight in $D^r[0, \infty)$ for each $N, X^{\epsilon}(\cdot)$ converges weakly to $X(\cdot)$, the unique solution to the martingale problem with initial condition $X(0) = X_0$.

Let \hat{C}_0 denote the space of real valued continuous functions on R^r , with compact support.

Theorem 2. For each $T < \infty$ and $N < \infty$ suppose that

$$(4.3) \quad \lim_{K \rightarrow \infty} \lim_{\epsilon > 0} P\{\sup_{t \leq T} |X^{\epsilon, N}(t)| \geq K\} = 0.$$

For each N and each $f(\cdot) \in \mathcal{D}_1$, a dense set in \hat{C}_0 , let $f^{\epsilon, N}(\cdot)$ be a sequence in $\mathcal{X}^{\epsilon, N}$ such that (4.4)-(4.6) hold.

$$(4.4) \quad \lim_{\epsilon \rightarrow 0} P\{\sup_{t \leq T} |f^{\epsilon, N}(t) - f(X^{\epsilon, N}(t))| \geq \alpha\} = 0, \text{ each } \alpha > 0, T < \infty.$$

For each $T < \infty$, let there be a random variable $M_T^{\epsilon, N}(f) \geq 0$ such that

$$(4.5) \quad \lim_{K \rightarrow \infty} \sup_{\epsilon > 0} P\{M_T^{\epsilon, N}(f) \geq K\} = 0,$$

where

$$(4.6) \quad \sup_{t \leq T} |\hat{A}^{\epsilon, N} f^{\epsilon, N}(t)| \leq M_T^{\epsilon, N}(f).$$

Then $\{f(X^{\epsilon, N}(\cdot))\}$, each $f(\cdot) \in \hat{C}_0$, and $\{X^{\epsilon, N}(\cdot)\}$ are tight in $D^T[0, \infty)$.

The usefulness and relative ease of application of Theorems 1 and 2 will become apparent in the following sections.

5. Assumptions for (3.1)

Two forms will be dealt with. The first allows \tilde{h}_ϵ , etc. to have a rather arbitrary dependence on ξ , but uses

(A1a) For each ϵ , $\{\xi_n^\epsilon\}$ is stationary, bounded, ϕ -mixing (see [7]) uniformly in ϵ , with mixing rate satisfying $\sum_i \phi_i^{1/2} < \infty$.

In many applications, it is desired to have $\{\xi_n^\epsilon\}$ unbounded (say Gaussian). Under additional restrictions on $\tilde{h}_\epsilon, g_\epsilon, k_\epsilon$, this can be treated for particular cases of $\{\xi_n^\epsilon\}$. We also treat the case, quite important in practice, where (A1b) holds in lieu of (A1a).

(A1b) Let the functions in (3.1) take the forms $g_\epsilon(x, \xi) = g_\epsilon(x)\xi$, $\tilde{h}_\epsilon(x, \xi) = h_\epsilon(x)\xi$, and for some $\alpha > 0$, $c > 0$, $k_\epsilon(x, \xi) = k_\epsilon(x)O(|\xi|^c + 1)$ and the $o(\epsilon)$ coefficient of k_ϵ is $\epsilon^{1+\alpha}$.
Let there be a matrix L and a vector valued stationary Markov process $\{\xi_n\}$ such that $E\xi_n = 0$, all moments are finite and $\xi_n^\epsilon \equiv L\xi_n$. Let $\{R_\ell\}$ be such that $R_\ell \geq |E\xi_n \xi_{n+\ell}'|$, $|E[\xi_\ell | \xi_0]| \leq R_\ell |\xi_0|$ and $\sum_\ell |R_\ell|^{1/2} < \infty$. Suppose that there are $\{\rho_\ell\}$ such that $\sum \rho_\ell^{1/2} < \infty$ and

$$(5.1) \quad |E(\xi_\ell \xi_\ell' | \xi_0) - E(\xi_\ell \xi_\ell')| \leq \rho_\ell (1 + |\xi_0|^2).$$

For each $\alpha > 0$, there is a $\beta > 0$ such that $E(|\xi_1|^\alpha | \xi_0) \leq (\text{constant})(|\xi_0|^\beta + 1)$.

Remarks on (A1a,b) follow (A8).

When Case (A1b) holds, the $k_\epsilon(x, \xi)$, etc., is to be replaced by $k_\epsilon(x)$, etc., in (A2) and (A5) below.

(A2) $k_\epsilon(\cdot, \cdot)$ is measurable, and uniformly bounded on bounded x-sets. $\tilde{h}_\epsilon(\cdot, \cdot), g_\epsilon(\cdot, \cdot)$ and $\bar{h}_\epsilon(\cdot)$ are measurable and continuous in x for each ξ . They are bounded on bounded x-sets uniformly in ϵ, ξ .

(A3) There is a continuous $\bar{h}(\cdot)$ such that $\bar{h}_\epsilon(\cdot) \rightarrow \bar{h}(\cdot)$ uniformly on bounded x-sets.

(A4) $E\tilde{h}_\epsilon(x, \xi_n^\epsilon) \equiv 0$, each n, ϵ, x .

(A5) $g_\epsilon(\cdot, \xi)$ and $\tilde{h}_\epsilon(\cdot, \xi)$ are (twice, once, resp.) continuously differentiable for each ξ, ϵ ; the derivatives are bounded on bounded x-sets, uniformly in ξ, ϵ .

(A6) There is a continuous function $\bar{G}_0(\cdot)$ such that

$$Eg_\epsilon(x, \xi_0^\epsilon)g'_\epsilon(x, \xi_0^\epsilon) \rightarrow \bar{G}_0(x)$$

uniformly on bounded x-sets. There are continuous (and symmetric w.l.o.g) $a_{ij}(\cdot)$ such that for each $f(\cdot, \cdot) \in \mathcal{S}_0^{1,2}$,

$$\begin{aligned} & \frac{1}{2} Eg'_\epsilon(x, \xi_0^\epsilon)f_{xx}(x, t)g_\epsilon(x, \xi_0^\epsilon) + \sum_{\ell=1}^{\infty} Eg'_\epsilon(x, \xi_\ell^\epsilon)f_{xx}(x, t)g_\epsilon(x, \xi_0^\epsilon) \\ & + \frac{1}{2} \sum_{i,j} a_{ij}(x)f_{x_i x_j}(x, t) \end{aligned}$$

uniformly on bounded x-sets. Thus, the second term on the left also

converges uniformly to the above right hand side

minus $\frac{1}{2}$ trace $G_0(x)f_{xx}(x,t)$.

(A7) There is a continuous $\bar{g}(\cdot)$ such that

$$\sum_{\ell=1}^{\infty} E g_{\epsilon, x}(x, \xi_{\ell}^{\epsilon}) g_{\epsilon}(x, \xi_0^{\epsilon}) \rightarrow \bar{g}(x)$$

uniformly on bounded x-sets as $\epsilon \rightarrow 0$, where

$$g_{\epsilon, x}(x, \xi) = \begin{bmatrix} g_{1, x_1}(x, \xi), \dots, g_{1, x_r}(x, \xi) \\ g_{r, x_1}(x, \xi), \dots, g_{r, x_r}(x, \xi) \end{bmatrix}$$

(A8) There is a unique solution (for each initial condition)

to the martingale problem in $D^r[0, \infty)$ with operator

$$A = \frac{1}{2} \sum_{i,j=1}^r a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^r [\bar{h}_i(x) + \bar{g}_i(x)] \frac{\partial}{\partial x_i},$$

where $\bar{h}_i(\cdot)$ and $\bar{g}_i(\cdot)$ are the i^{th} components of the vectors $\bar{h}(\cdot)$ and $\bar{g}(\cdot)$, resp.

Remarks on (A1a,b). We note first that (5.1) holds if $\{\xi_n\}$ is a zero mean stationary Gaussian process with correlation $E\xi_0 \xi_{\ell}' = R_{\ell}$ and $\sum_{\ell} |R_{\ell}|^{1/2} < \infty$. For notational convenience, we show it in the unit variance and scalar case. We can write $\xi_{\ell} = R_{\ell} \xi_0 + \tilde{\xi}_{\ell}$, where ξ_0 and $\tilde{\xi}_{\ell}$ are zero mean and independent. The sentence below (5.1) obviously holds here. Also, equation (5.1) is implied by the calculations $E\xi_{\ell}^2 = 1 = R_{\ell}^2 + E\tilde{\xi}_{\ell}^2$,

$$E(\xi_k^2 | \xi_0) - E\xi_k^2 = R_k^2 \xi_0^2 + E\xi_k^2 - 1.$$

It can be seen from the method of proof that the theorem would remain valid under weaker conditions than (A1a) (if the case (A1b) does not hold). Then the special structure of the functions in (3.1) would need to be taken into account. The proof is given for the broad and standard cases (A1a,b), but its general outline would be followed in other cases. For example, there are important examples in communication theory (for example) where a system processes some wide-band and unbounded noise input non-linearly. Somewhat analogous "continuous parameters" techniques have been developed [8] to deal with a number of these important but special cases. Presumably, similar results are possible in the discrete parameter case.

In the state-dependent noise case of Section 9, (A1) is not used, and is replaced by a longer list of the (weaker) specific conditions which are actually used in the proof (of either Theorems 3 or 4). See also the comments on the Markov case in Section 9, where the conditions on the continuity and differentiability of the functions are weakened. We note also that the proofs can readily be adapted to the traditional stochastic approximation case where ϵ is replaced by a sequence $\{\epsilon_n\}$, with $\epsilon_n > 0$ used at iterate n , and $\sum \epsilon_n = \infty$, $\epsilon_n \rightarrow 0$.

6. The Main Convergence Theorem

The same $\{f^{\epsilon,N}(\cdot)\}$ will be used to satisfy the requirements of both Theorems 1 and 2. We use $\mathcal{B} = \mathcal{B}_0^{1,3}$ and, given $f(\cdot, \cdot) \in \mathcal{B}$,

construct $f^{\epsilon, N}(\cdot)$ via a discrete parameter analog of the method used in [5], [6], [9]. The $\{X_n^\epsilon\}$ and $\{X_n^{\epsilon, N}\}$ are defined by (3.1) and (3.2) resp., here.

Theorem 3. Let $X_0^\epsilon \rightarrow X_0$ weakly. Under (A1a or b) and (A2)-(A8), $\{X^\epsilon(\cdot)\}$ is tight in $D^r[0, \infty)$ and, as $\epsilon \rightarrow 0$, converges weakly to the diffusion determined by the solution to the martingale problem with the operator A of (A8), and initial condition $X_0 = X(0)$.

Proof. Part 1. Fix $f(\cdot, \cdot) \in \mathcal{D}$ and fix N throughout the proof.

First $\{f^{\epsilon, N}(\cdot)\}$ satisfying (4.1), (4.2) will be found. In order to simplify the notation in the proof the superscript N on $\hat{A}^{\epsilon, N}$ and on $X_n^{\epsilon, N}$, $X^{\epsilon, N}(\cdot)$ and $X^N(\cdot)$ and the subscript N on $h_{\epsilon, N}$, etc., will henceforth be dropped (in the proof only), but we always work here with $X_n^{\epsilon, N}$, $h_{\epsilon, N}$, etc., only, and not with the original X_n and h_{ϵ} , etc.

The proof under (Ala) will be given first. The simple modifications required under (Alb) will then be stated.

Thus we can suppose that there is a constant K_N such that $|X_n^{\epsilon}| \leq K_N$, $|X_{n+1}^{\epsilon} - X_n^{\epsilon}| \leq K_N \sqrt{\epsilon}$ and that $|\tilde{h}_{\epsilon}|$, $|\bar{h}_{\epsilon}|$, etc., (definitions (Ala or Alb) are bounded above by K_N .

The $O(\cdot)$, $o(\cdot)$ and $o_i(\cdot)$ terms (with or without subscripts) are uniform in all variables (except N , which is fixed throughout) other than their arguments, unless otherwise stated. Evaluate

$$\begin{aligned}
 \epsilon \hat{A}^{\epsilon} f(X_n^{\epsilon}, n\epsilon) &= E_n^{\epsilon} f(X_{n+1}^{\epsilon}, n\epsilon + \epsilon) - f(X_n^{\epsilon}, n\epsilon) \\
 &= E_n^{\epsilon} [f(X_{n+1}^{\epsilon}, n\epsilon + \epsilon) - f(X_{n+1}^{\epsilon}, n\epsilon)] \\
 &\quad + E_n^{\epsilon} [f(X_{n+1}^{\epsilon}, n\epsilon) - f(X_n^{\epsilon}, n\epsilon)] \\
 &= \epsilon E_n^{\epsilon} f_t(X_{n+1}^{\epsilon}, n\epsilon) + o(\epsilon) + f'_x(X_n^{\epsilon}, n\epsilon) E_n^{\epsilon} [\epsilon \bar{h}_{\epsilon}(X_n^{\epsilon}) \\
 &\quad + \sqrt{\epsilon} g_{\epsilon}(X_n^{\epsilon}, \xi_n^{\epsilon}) + \epsilon \tilde{h}(X_n^{\epsilon}, \xi_n^{\epsilon})] \\
 &\quad + \frac{\epsilon}{2} E_n^{\epsilon} g'_{\epsilon}(X_n^{\epsilon}, \xi_n^{\epsilon}) f'_{xx}(X_n^{\epsilon}, n\epsilon) g_{\epsilon}(X_n^{\epsilon}, \xi_n^{\epsilon}) + o_1(\epsilon).
 \end{aligned}
 \tag{6.1}$$

Part 2. Define $f^\epsilon = f(X_n^\epsilon, n\epsilon) + f_1^\epsilon(n\epsilon) + f_2^\epsilon(n\epsilon)$, where the $f_1^\epsilon(\cdot)$ will be selected so that (4.1) and (4.2) hold. We now define $f_1^\epsilon(\cdot)$ in such a way that we "average out" the ϵ_n^ϵ dependent parts of (6.1). Define $G_\epsilon(x, \xi) = g_\epsilon(x, \xi)g'_\epsilon(x, \xi)$ and $\bar{G}_\epsilon(x) = E G_\epsilon(x, \xi)$. Define $f_1(t)$ to be constant on each $[n\epsilon, n\epsilon + \epsilon)$ and satisfy $f_1^\epsilon(n\epsilon) = f_1^\epsilon(X_n^\epsilon, n\epsilon)$, where $f_1^\epsilon(\cdot, \cdot)$ is defined by

$$\begin{aligned} f_1^\epsilon(x, n\epsilon) = & \epsilon \sum_{\ell=n}^{\infty} E_n^\epsilon f'_x(x, \epsilon n) \tilde{h}_\epsilon(x, \xi_\ell^\epsilon) \\ & + \frac{\epsilon}{2} \sum_{\ell=n}^{\infty} [E_n^\epsilon \text{trace } G_\epsilon(x, \xi_\ell^\epsilon) \cdot f_{xx}(x, n\epsilon) - \\ & \quad - \text{trace } \bar{G}_\epsilon(x) f_{xx}(x, n\epsilon)] \\ & + \sqrt{\epsilon} \sum_{\ell=n}^{\infty} E_n^\epsilon f'_x(x, \epsilon n) g_\epsilon(x, \xi_\ell^\epsilon) = T_1^\epsilon + T_2^\epsilon + T_3^\epsilon. \end{aligned}$$

By the ϕ -mixing and the facts that the expectations of all the summands are zero, the first two terms are $O(\epsilon)$, and the third $O(\sqrt{\epsilon})$, uniformly in x . Thus, $p\text{-}\lim_{\epsilon \rightarrow 0} f_1^\epsilon(\cdot) = 0$.

Now calculate $\hat{A}^\epsilon f_1^\epsilon(\cdot)$ at $t = n\epsilon$:

$$\epsilon \hat{A}^\epsilon f_1^\epsilon(n\epsilon) = -\epsilon E_n^\epsilon f'_x(X_n^\epsilon, \epsilon n) \tilde{h}_\epsilon(X_n^\epsilon, \xi_n^\epsilon)$$

$$- \frac{\epsilon}{2} E_n^\epsilon \text{trace } G_\epsilon(X_n^\epsilon, \xi_n^\epsilon) f_{xx}(X_n^\epsilon, n\epsilon) +$$

$$+ \frac{\epsilon}{2} \text{trace } \bar{G}_\epsilon(X_n^\epsilon) f_{xx}(X_n^\epsilon, n\epsilon)$$

$$- \sqrt{\epsilon} E_n^\epsilon f'_x(X_n^\epsilon, \epsilon n) g_\epsilon(X_n^\epsilon, \xi_n^\epsilon)$$

$$(6.2) \quad + \epsilon \sum_{\ell=n+1}^{\infty} E_n^\epsilon [f'_x(X_{n+1}^\epsilon, \epsilon n + \epsilon) \tilde{h}_\epsilon(X_{n+1}^\epsilon, \xi_\ell^\epsilon) -$$

$$- f'_x(X_n^\epsilon, \epsilon n) \tilde{h}_\epsilon(X_n^\epsilon, \xi_\ell^\epsilon)]$$

$$+ [\text{a term similar to the last but coming from } T_2^\epsilon \text{ instead of from } T_1^\epsilon] +$$

$$+ \sqrt{\epsilon} \sum_{\ell=n+1}^{\infty} E_n^\epsilon [f'_x(X_{n+1}^\epsilon, \epsilon n + n) g_\epsilon(X_{n+1}^\epsilon, \xi_\ell^\epsilon)$$

$$- f'_x(X_n^\epsilon, \epsilon n) g_\epsilon(X_n^\epsilon, \xi_\ell^\epsilon)].$$

The first, second and fourth terms in (6.2) cancel terms in

(6.1) (which is the reason for constructing $f_1^\epsilon(\cdot)$ as we did).

The 5th and 6th terms are $o(\epsilon)$, as will now be shown. It is easy to see that the difference between these terms and their values with $\epsilon n + \epsilon$ replaced by ϵn in the $f_x(\cdot, \cdot)$ is $o(\epsilon)$.

The rest will be shown for the 5th term only. Define $H_\epsilon(x, \epsilon n, \xi) = f'_x(x, \epsilon n) \tilde{h}_\epsilon(x, \xi)$. Then we must show that

$$(6.3) \quad \epsilon \sum_{\ell=n+1}^{\infty} E_n^\epsilon [H_\epsilon(X_{n+1}^\epsilon, \epsilon n, \xi_\ell^\epsilon) - H_\epsilon(X_n^\epsilon, \epsilon n, \xi_\ell^\epsilon)] = o(\epsilon).$$

By (the zeroeth order) Taylor's formula with remainder, (6.3) equals

$$\epsilon \sum_{\ell=n+1}^{\infty} \int_0^1 E_n^{\epsilon} H'_{\epsilon, x}(X_n^{\epsilon} + s(X_{n+1}^{\epsilon} - X_n^{\epsilon}), \epsilon n, \xi_{\ell}^{\epsilon}) [X_{n+1}^{\epsilon} - X_n^{\epsilon}] ds$$

which equals $O(\epsilon^{3/2})$ by the ϕ -mixing condition and the fact that $|X_{n+1}^{\epsilon} - X_n^{\epsilon}| \leq K_N \sqrt{\epsilon}$. The 6th term is treated similarly. Thus,

$$p\text{-}\lim_{\epsilon \rightarrow 0} (5^{\text{th}} + 6^{\text{th}} \text{ terms})/\epsilon = 0.$$

The 7th and last term remains. Again, the 7th term differs by $o(\epsilon)$ from the term obtained by replacing $\epsilon n + \epsilon$ by ϵn in the $f_x(\cdot, \cdot)$, and we replace $\epsilon n + \epsilon$ by ϵn there. Write $K_{\epsilon}(x, \epsilon n, \xi) = f'_x(x, \epsilon n) g_{\epsilon}(x, \xi)$. Then by applying (the first order) Taylor's formula with remainder, we can write the 7th term as

$$(6.4) \quad o(\epsilon) + \sqrt{\epsilon} \sum_{\ell=n+1}^{\infty} E_n^{\epsilon} [K'_{\epsilon, x}(X_n^{\epsilon}, \epsilon n, \xi_{\ell}^{\epsilon}) (X_{n+1}^{\epsilon} - X_n^{\epsilon}) + \int_0^1 (1-s) (X_{n+1}^{\epsilon} - X_n^{\epsilon})' K_{\epsilon, xx}(X_n^{\epsilon} + s(X_{n+1}^{\epsilon} - X_n^{\epsilon}), \epsilon n, \xi_{\ell}^{\epsilon}) (X_{n+1}^{\epsilon} - X_n^{\epsilon}) ds]$$

The contribution of the "sum of the integrals" component of (6.4) is $O(\epsilon^{3/2})$ by the ϕ -mixing and the fact that $E K_{\epsilon, xx}(x, \epsilon n, \xi_{\ell}^{\epsilon}) = 0$.

Next, by collecting the terms of the first component of the sum in (6.4) according to their power of ϵ , we can write

$$(6.5) \quad (6.4) = o(\epsilon) + \epsilon \sum_{\ell=n+1}^{\infty} E_n^{\epsilon} [f'_x(X_n^{\epsilon}, \epsilon n) g_{\epsilon}(X_n^{\epsilon}, \xi_{\ell}^{\epsilon})]' g_{\epsilon}(X_n^{\epsilon}, \xi_{\ell}^{\epsilon}).$$

Part 3. Next, $f_2(\cdot)$ will be introduced in a way which will "average out" the second term on the right side of (6.5). Note that, as $\epsilon \rightarrow 0$ (and $n\epsilon \rightarrow t$ in $f(\cdot, \cdot)$)

$$\sum_{\ell=1}^{\infty} E[f'_x(x, n\epsilon) g_\epsilon(x, \xi_\ell^\epsilon)]'_x g_\epsilon(x, \xi_0^\epsilon)$$

converges uniformly to the quantity mentioned in the last sentence of (A6) plus $f'_x(x, t) \bar{g}(x)$ ($\bar{g}(\cdot)$ is defined in (A7)). Recall that we have suppressed the affixes N , and that the $g_\epsilon(x, \xi)$, etc. used in the proof is actually (in terms of the original $g_\epsilon(x, \xi)$, etc., $g_\epsilon(x, \xi) q_N(x)$, etc.

Next, $f_2^\epsilon(\cdot)$ is to be chosen. It is to be constant on each interval $[n\epsilon, n\epsilon + \epsilon)$ and satisfy $f_2^\epsilon(n\epsilon) = f_2^\epsilon(X_n^\epsilon, n\epsilon)$, where $f_2^\epsilon(x, n\epsilon)$ is defined by

$$f_2^\epsilon(n\epsilon) = f_2^\epsilon(X_n^\epsilon, n\epsilon) \text{ where } f_2^\epsilon(x, n\epsilon) \text{ is defined by}$$

$$f_2^\epsilon(x, n\epsilon) = \epsilon \sum_{\ell=n}^{\infty} \sum_{j=\ell+1}^{\infty} [E_n^\epsilon(f'_x(x, n\epsilon) g_\epsilon(x, \xi_j^\epsilon))'_x g_\epsilon(x, \xi_\ell^\epsilon) - E(f'_x(x, n\epsilon) g_\epsilon(x, \xi_j^\epsilon))'_x g_\epsilon(x, \xi_\ell^\epsilon)].$$

By the ϕ -mixing and the fact that the centered summand has zero expectation, $|f_2^\epsilon(x, n\epsilon)| = O(\epsilon)$ uniformly in x . Thus, $p\text{-}\lim_{\epsilon \rightarrow 0} f_2^\epsilon(\cdot) = 0$. Next, evaluating $\hat{A}^\epsilon f_2^\epsilon(n\epsilon)$ and using the stationarity of the $\{\xi_j^\epsilon\}$,

$$\begin{aligned} \hat{\epsilon} \hat{A}^\epsilon f_2^\epsilon(n\epsilon) = & -\epsilon \sum_{\ell=n+1}^{\infty} E_n^\epsilon(f'_X(X_n^\epsilon, n\epsilon) g_\epsilon(X_n^\epsilon, \xi_\ell^\epsilon))' g_\epsilon(X_n^\epsilon, \xi_n^\epsilon) \\ & + \epsilon \sum_{\ell=1}^{\infty} E(f'_X(x, n\epsilon) g_\epsilon(x, \xi_\ell^\epsilon))' g_\epsilon(x, \xi_0^\epsilon) \Big|_{x=X_n^\epsilon} \end{aligned}$$

(6.6)

$$\begin{aligned} & + \epsilon \sum_{n+1}^{\infty} \sum_{j=\ell+1}^{\infty} E_n^\epsilon[Q_\epsilon(X_{n+1}^\epsilon, \epsilon+n\epsilon, \xi_j^\epsilon, \xi_\ell^\epsilon) \\ & - Q_\epsilon(X_n^\epsilon, n\epsilon, \xi_j^\epsilon, \xi_\ell^\epsilon)], \end{aligned}$$

where

$$\begin{aligned} Q_\epsilon(x, n\epsilon, \xi_j^\epsilon, \xi_\ell^\epsilon) = & (f'_X(x, n\epsilon) g_\epsilon(x, \xi_j^\epsilon))' g_\epsilon(x, \xi_\ell^\epsilon) - \\ & - E(f'_X(x, n\epsilon) g_\epsilon(x, \xi_j^\epsilon))' g_\epsilon(x, \xi_\ell^\epsilon). \end{aligned}$$

By (A1a) and an argument using Taylor's formula similar to that used in connection with (6.3), we get that the last term of (6.6) is $o(\epsilon)$. Also, the first term of (6.6) cancels the second term on the right of (6.4).

Part 4. Recall the definition of $f^\epsilon(\epsilon n)$ given in Part 2.

Adding the results of Parts 1-3, and deleting the terms of $\hat{A}^\epsilon f^\epsilon(\cdot)$ which cancel, yields (modulo the $q_N(\cdot)$ factors of \bar{H}_ϵ , etc.)

$$f^\epsilon(\epsilon n) = f(X_n^\epsilon, n\epsilon) + O(\sqrt{\epsilon})$$

$$\hat{A}^\epsilon f^\epsilon(\epsilon n) = f_t(X_n^\epsilon, n\epsilon) + f'_x(X_n^\epsilon, n\epsilon) \bar{h}_\epsilon(X_n^\epsilon)$$

$$+ \frac{1}{2} E g'_\epsilon(x, \xi_n^\epsilon) f_{xx}(x, n\epsilon) g_\epsilon(x, \xi_n^\epsilon) \Big|_{x=X_n^\epsilon}$$

$$+ \sum_{l=1}^{\infty} E(f'_x(x, n\epsilon) g_\epsilon(x, \xi_l^\epsilon))' g_\epsilon(x, \xi_0^\epsilon) \Big|_{x=X_n^\epsilon} + o(\epsilon)/\epsilon.$$

Then, by the convergences (A3), (A6), and (A7) all of (4.4), (4.5) and (4.6) hold (the $M_T^{\epsilon, N}(f)$ are bounded here), yielding tightness; also (4.1), (4.2) hold, yielding the asserted weak convergence.

Part 5. Owing to the special form of the functions under (A1b), essentially the same proof can be used. The only problem is that the $O(\cdot)$ and $o(\cdot)$ terms are not uniform in $\bar{\xi}_{n-1}$. Let us examine some typical terms. Let E_n denote conditioning on $\bar{\xi}_i, i < n$. All the $O(\cdot)$ below are uniform in all variables other than their argument. The essential part of the last term in the definition of $f_1^\epsilon(\epsilon n)$ is (the f_x term is omitted)

$$|\sqrt{\epsilon} \sum_{\ell=n}^{\infty} E_n^\epsilon g_\epsilon(x, \xi_\ell^\epsilon)| = |\sqrt{\epsilon} g_\epsilon(x) \sum_{\ell=n}^{\infty} L E_n^\epsilon \xi_\ell^\epsilon| \leq$$

$$\sqrt{\epsilon} |g_\epsilon(x) L| \sum_{\ell=n}^{\infty} R_{\ell-n+1} |\bar{\xi}_{n-1}^\epsilon|.$$

Next, examine the expression below (6.3). For some $c_1 > 0$, we can show that it is bounded above by

$$O(\epsilon^{3/2}) \sum_{\ell=n+1}^{\infty} E_n^\epsilon |E_{n+1}^\epsilon \bar{\xi}_\ell| (1 + |\bar{\xi}_n|^{c_1})$$

$$= O(\epsilon^{3/2}) E_n^\epsilon |\bar{\xi}_n| (1 + |\bar{\xi}_n|^{c_1}) \sum_{\ell} |R_\ell|.$$

Also, by a careful use of iterated conditional expectations and (5.1), we can show that

$$|f_2^\epsilon(n\epsilon)| \leq O(\epsilon)(1+|\bar{\xi}_{n-1}|^2).$$

In fact, it can readily be shown that $|f_i^\epsilon(n\epsilon)|$ and all the $o(\epsilon)/\epsilon$ terms in $\hat{A}^\epsilon f^\epsilon(n\epsilon)$ which appeared in Parts 1 to 4 are of the form $\epsilon^\gamma(1+|\bar{\xi}_{n-1}|^\delta)$ for some $\gamma > 0$ and $\delta \geq 0$. This, together with the fact that all moments exist implies (4.1)-(4.2).

Next, note that for each $m \geq 1$,

$$\begin{aligned} P\left\{\sup_{n \leq T/\epsilon} \epsilon^\gamma |\bar{\xi}_n|^\delta \geq u\right\} &\leq \frac{T}{\epsilon} P\{\epsilon^\gamma |\bar{\xi}_n|^\delta \geq u\} \\ &\leq \frac{TE|\bar{\xi}_n|^{\delta m} \epsilon^{\gamma m}}{\epsilon u^m}. \end{aligned}$$

By letting $\gamma m > 1$, (4.4) to (4.6) hold. Q.E.D.

7. Application to a Problem of Guess and Gillespie [1]

In [1], the scalar problem (7.1) is treated

$$(7.1) \quad x_{n+1}^\epsilon = x_n^\epsilon + f(S_n^\epsilon) + (\exp g(S_n^\epsilon) - 1)x_n^\epsilon, \quad x_0^\epsilon = x_0,$$

under two different sets of conditions. The first being that for some constants $\mu, \sigma^2 < 0$ and $\{r_n\}$ such that $\sum_n r_n < \infty$,

$ES_n^\epsilon = \mu\epsilon$, $\text{var } S_n^\epsilon = \sigma^2\epsilon$, $\text{Cov}(S_n^\epsilon, S_0^\epsilon) = \epsilon\sigma^2 r_n$ and $\{(S_n^\epsilon - ES_n^\epsilon)/\sqrt{\epsilon}\}$ satisfying the condition on $\{\xi_n^\epsilon\}$ in (A1a). Under their second set of conditions the $\{S_n^\epsilon\}$ are Gaussian, hence unbounded. Our method works in either case. We stick to the first case. Note that $|S_n^\epsilon| \leq K\sqrt{\epsilon}$ for some real K . Let $f(\cdot), g(\cdot)$ have continuous third derivatives with $f(0) = g(0) = 0$, and put (7.1) into the form (3.1) by expanding $\exp g(\cdot)$ and $f(\cdot)$. First write

$$(7.2) \quad X_{n+1}^\epsilon = X_n^\epsilon + \epsilon f(S_n^\epsilon) + E(\exp g(S_{n+1}^\epsilon) - 1)X_n^\epsilon + \\ + \sqrt{\epsilon} \frac{(f(S_n^\epsilon) - Ef(S_n^\epsilon))}{\sqrt{\epsilon}} + \\ + \sqrt{\epsilon} \left[\frac{\exp(g(S_{n+1}^\epsilon) - 1) - E(\exp g(S_{n+1}^\epsilon) - 1)}{\sqrt{\epsilon}} \right] X_n^\epsilon.$$

Expanding $g(\cdot)$ and $f(\cdot)$ yields (the $o(\cdot)$ are uniform in all variables but their argument).

$$(7.3) \quad X_{n+1}^\epsilon = X_n^\epsilon + \epsilon[f_s(0)\mu + \frac{1}{2} f_{ss}(0)\sigma^2 + \mu g_s(0)]X_n^\epsilon +$$

$$\frac{1}{2} (g_{ss}(0) + g_s^2(0))\sigma^2 X_n^\epsilon + \\ + \sqrt{\epsilon} \left[f_s(0) \frac{(S_n^\epsilon - ES_n^\epsilon)}{\sqrt{\epsilon}} + g_s(0) \frac{(S_n^\epsilon - ES_n^\epsilon)X_n^\epsilon}{\sqrt{\epsilon}} \right] + \\ + \frac{\epsilon}{2} \left[f_{ss}(0) \frac{((S_n^\epsilon)^2 - E(S_n^\epsilon)^2)}{\epsilon} + \right. \\ \left. + (g_{ss}(0) + g_s^2(0)) \frac{((S_n^\epsilon)^2 - E(S_n^\epsilon)^2)X_n^\epsilon}{\epsilon} \right] + o(\epsilon).$$

$$\equiv X_n^\epsilon + \epsilon \bar{h}(X_n^\epsilon) + \sqrt{\epsilon} g_\epsilon(X_n^\epsilon, \xi_n^\epsilon) + \epsilon \tilde{h}_\epsilon(X_n^\epsilon, \xi_n^\epsilon) + o(\epsilon),$$

where ξ_n^ϵ , $\bar{h}(\cdot)$, $g_\epsilon(\cdot)$ and $\tilde{h}_\epsilon(\cdot)$ are defined in the obvious manner.

Equation (7.3) is of the form (3.1) and the conditions of Theorem 3 hold. An immediate application of Theorem 3 yields that the $\{X^\epsilon(\cdot)\}$ are tight and that the weak limit is the unique diffusion with the operator (the same result as [1])

$$A = [a_1 + a_2 x] \frac{\partial}{\partial x} + \frac{1}{2} (b_1 + b_2 x)^2 \frac{\partial^2}{\partial x^2}$$

$$a_1 = \mu f_s(0) + \frac{1}{2} \sigma^2 [(\eta^2 - 1) f_s(0) g_s(0) + f_{ss}(0)]$$

$$a_2 = \mu g_s(0) + \frac{1}{2} \sigma^2 [\eta^2 g_s^2(0) + g_{ss}(0)]$$

$$b_1 = \sigma \eta f_s(0), \quad b_2 = \sigma \eta g_s(0)$$

$$\eta^2 = 1 + 2 \sum_{i=1}^{\infty} r_i.$$

Since our method does not require an explicit construction of the solution process (as essentially done in the proof in [1]), it is equally applicable to vector-valued versions of (7.1), and seems to give a bit more insight into the approximation process and the effects of variations in the data.

8. Rate of Convergence for a Stochastic Approximation

Stochastic approximations of the form

$$(8.1) \quad X_{n+1}^\epsilon = X_n^\epsilon + \epsilon h(X_n^\epsilon) + \epsilon g(X_n^\epsilon, \xi_n^\epsilon)$$

have numerous applications in sequential monte-carlo optimization in control theory. In [10], [11], convergence and rate of convergence is treated via a rather direct method. We use the notation of this paper rather than that of [10], [11]. Let $h(\cdot)$ and $g(\cdot, \cdot)$ satisfy (A5) and assume $Eg(x, \xi_n^\epsilon) \equiv 0$. Let $x(t) \equiv \theta$ denote a globally asymptotically stable solution to $\dot{x} = h(x)$ (which we assume exists). Thus, $h(\theta) = 0$. The properties of $\{X_n^\epsilon\}$ for large n and small ϵ are of interest. To investigate them, define $U_n^\epsilon = (X_n^\epsilon - \theta)/\sqrt{\epsilon}$. In [11], it is shown under appropriate conditions that there is an $N_\epsilon \rightarrow \infty$ as $\epsilon \rightarrow 0$ (unless $g(x, \xi) = g(x)\xi$, in which case the process can be centered so that $N_\epsilon \equiv 0$) such that $\{U_n^\epsilon, n \geq N_\epsilon, \epsilon > 0\}$ is tight. Let us assume this tightness here and study the asymptotics of $U^\epsilon(\cdot)$ where $U^\epsilon(t) = U_{N_\epsilon + n}^\epsilon$ on $[n\epsilon, n\epsilon + \epsilon)$. The "tail" of $U^\epsilon(\cdot)$ for small ϵ contains the rate of convergence information for $\{X_n^\epsilon\}$, for small ϵ . By the assumed tightness

$$\lim_{\epsilon \rightarrow 0} \overline{\lim}_{n \rightarrow \infty} P\{|X_n^\epsilon - \theta| \geq \delta\} = 0, \text{ each } \delta > 0.$$

Now, by (8.1) we can write

$$(8.2) \quad U_{n+1}^\epsilon = U_n^\epsilon + \epsilon[h_x(\theta) + g_x(\theta, \xi_n^\epsilon)]U_n^\epsilon + \sqrt{\epsilon} g(\theta, \xi_n^\epsilon) + \epsilon^{3/2} B_0(G(X_n^+), U_n^\epsilon)U_n^\epsilon,$$

where $B_0(\cdot, \cdot)$ is a matrix valued bilinear form, $G(\cdot)$ is a function which is bounded on bounded x -sets and $X_n^+ \in [X_n^\epsilon, X_{n+1}^\epsilon]$. Next fix a weakly convergent subsequence of $\{U_{N_\epsilon}^\epsilon\}$ with limit U_0 . Then under

suitable conditions on $\{g(\theta, \xi_n^\epsilon)\}$, Theorem 3 and the truncation method can be applied to get Theorem 2 of [10]; in particular, $U(\cdot)$, the weak limit of $\{U^\epsilon(\cdot)\}$ solves

$$(8.3) \quad dU = h_x(\theta)Udt + R^{1/2}dB, \quad U(0) = U_0$$

where $R = \lim_{\epsilon \rightarrow 0} \sum_{i=-\infty}^{\infty} E g(\theta, \xi_0^\epsilon) g'(\theta, \xi_i^\epsilon)$ and $B(\cdot)$ is a standard Wiener process. This is one of the results of [11] and was proved there by solving (8.2) (modulo the B_0 -terms) and essentially constructing the limit via weak convergence theory, a more arduous task than merely using Theorem 3 (although the proof in [11] has its own intrinsic interest, and the paper contains other interesting results).

If N_ϵ is chosen such that $\epsilon N_\epsilon \rightarrow \infty$ as $\epsilon \rightarrow 0$, then the weak limit $U(\cdot)$ of the original $\{U^\epsilon(\cdot)\}$ is the stationary solution to (8.3). The technique can also be applied to the rate of convergence problem for stochastic approximations where the ϵ in (8.1) is replaced by ϵ_n and $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$, with $\sum \epsilon_n = \infty$, although we will not pursue this. In this case, the interpolation intervals would be ϵ_n , rather than the constant ϵ .

9. State Dependent Noise $\{\xi_n^\epsilon\}$

Under broad conditions, the treatment for state dependent noise is very similar to that given in Theorem 3, and only two cases will be discussed. For each $\epsilon > 0$, $\{\psi_n^\epsilon\}$ denotes a stationary bounded sequence. If the $\{\psi_n^\epsilon\}$ are mutually independent, then the

$\{\xi_{n-1}^\epsilon, X_n^\epsilon\}$ below are Markovian and, as will be seen, the smoothness of Q_ϵ in (A9) below can then be weakened. The scheme introduced here covers some interesting cases, but it should be seen as being typical of the possibilities, as indicated by the example in Section 10, which uses a slightly different setup. We will need:

(A9) There is a function $Q_\epsilon(\cdot, \cdot, \cdot)$ such that $\xi_\ell^\epsilon = Q_\epsilon(X_\ell^\epsilon, \xi_{\ell-1}^\epsilon, \psi_\ell^\epsilon)$ where Q has continuous (uniformly in ψ -in bounded (x, ξ) sets) partial (x, ξ) derivatives up to second order. For each sphere S_N there is a sphere $S_{N, \xi}$ such that ξ_ℓ^ϵ remains in $S_{N, \xi}$ as long as X_ℓ^ϵ remains in S_N .

It is now convenient to introduce some auxiliary processes which will be used in the construction of the $f_1^\epsilon(\cdot)$. For each n , define the processes $P_n(x) = \{\xi_\ell^\epsilon(x), \xi_{\ell, x}^\epsilon(x), \xi_{\ell, xx}^\epsilon(x), \ell \geq n\}$ as follows:

$$\begin{aligned} \xi_\ell^\epsilon(x) &= Q_\epsilon(x, \xi_{\ell-1}^\epsilon(x), \psi_\ell^\epsilon) \\ (9.1) \quad \xi_{\ell, x}^\epsilon(x) &= Q_{\epsilon, x}(x, \xi_{\ell-1}^\epsilon(x), \psi_\ell^\epsilon) + Q_{\epsilon, \xi}(x, \xi_{\ell-1}^\epsilon(x), \psi_\ell^\epsilon) \xi_{\ell-1, x}^\epsilon(x), \end{aligned}$$

$$\xi_{\ell, xx}^\epsilon(x) = \text{collection of second partial derivatives of the components of } \xi_\ell^\epsilon(x).$$

The initial conditions on $P_n(x)$ will be given below, when f_1^ϵ and f_2^ϵ are defined.

(A10) For each x , there is a unique stationary process
 $\bar{P}(x) = \{\bar{\xi}_l^\epsilon(x), \bar{\xi}_{l,x}^\epsilon(x), \bar{\xi}_{l,xx}^\epsilon(x), \infty > l > -\infty\}$ satisfying (9.1).
The process is bounded (uniformly in bounded x -sets), and the
second (resp., third) member is the x -derivative of the first (resp.,
second) member.

Some sort of mixing condition on $P_n(x)$ is needed, as is some condition on the rate of convergence of the distributions of the "tails" of $P_n(x)$ to those of the stationary process $\bar{P}(x)$. A general type of mixing condition (as in (A1)) can be used, but we prefer here to introduce the conditions in the weakest and most explicit form (A11), (A12) that we can use. This is because, owing to the double requirement just mentioned, it seems simpler to do it this way. There are many sets of sufficient conditions which imply (A11)-(A12), and there seems little point in restricting the applicability of the result. Note that (A12) is the weakest condition which is usable in Theorem 3 in place of (A1) in the non-state-dependent noise case. Below, the subscript x denotes a total derivative; i.e., for a smooth function $q(\cdot)$,

$$q(x, \xi_l^\epsilon(x))_x = q_x(x, \xi_l^\epsilon(x)) + q_{\xi}(x, \xi_l^\epsilon(x)) \xi_{l,x}^\epsilon(x).$$

Conditions (A11) and (A12) can be combined, but it is perhaps more natural (for purposes of verification) to keep them as they are.

(A11) concerns the rate of convergence of the distributions of $P_n(x)$ to those of $\bar{P}(x)$ for "non-stationary" initial conditions for $P_n(x)$, while (A12) is more of a mixing condition, concerning the rate of convergence $[E_n^{\epsilon} q(\xi_{\ell+n}(x), \xi_{\ell+n,x}(x), \xi_{\ell+n,xx}(x)) - Eq(\xi_{\ell+n}(x), \xi_{\ell+n,x}(x), \xi_{\ell+n,xx}(x))] \rightarrow 0$ for smooth $q(\cdot)$, as $\ell \rightarrow \infty$.

(A11) For each $f(\cdot, \cdot) \in \mathcal{B}_0^{\infty, \infty}$, the sums (9.2), (9.3) converge (absolutely) uniformly in x, t, n and in the initial conditions $\xi_n^{\epsilon}(x), \xi_{n,x}^{\epsilon}(x)$ in bounded sets. Define

$$H_l^\epsilon(x) = E g'_\epsilon(x, \xi_l^\epsilon(x)) f_{xx}(x, t) g_\epsilon(x, \xi_l^\epsilon(x)) \\ - E g'_\epsilon(x, \bar{\xi}_l^\epsilon(x)) f_{xx}(x, t) g_\epsilon(x, \bar{\xi}_l^\epsilon(x))$$

$$I_{j,l}^\epsilon(x) = E(g'_\epsilon(x, \xi_j^\epsilon(x)) f_x(x, t))'_x g_\epsilon(x, \xi_l^\epsilon(x)) \\ - E(g'_\epsilon(x, \bar{\xi}_j^\epsilon(x)) f_x(x, t))'_x g_\epsilon(x, \bar{\xi}_l^\epsilon(x)).$$

$$(9.2) \quad \sum_{l=n}^{\infty} H_l^\epsilon(x), \quad \sum_{l=n}^{\infty} (H_l^\epsilon(x))_x$$

$$(9.3) \quad \sum_{l=n}^{\infty} \sum_{j=l+1}^{\infty} I_{j,l}^\epsilon(x), \quad \sum_{l=n}^{\infty} \sum_{j=l+1}^{\infty} (I_{j,l}^\epsilon(x))_x$$

Let E_n^ϵ denote conditioning on ψ_i^ϵ , $i \leq n$, and on the initial condition of $P_n(x)$, namely, $(\xi_n^\epsilon(x), \xi_{n,x}^\epsilon(x), \xi_{n,xx}^\epsilon(x))$. Owing to the stationarity of $\{\psi_l^\epsilon\}$ we could set $n \equiv 0$ in (A11).

(A12) Define $(j > l \geq n)$

$$J_l^\epsilon(x) = E_n^\epsilon f'_x(x, t) \tilde{h}_\epsilon(x, \xi_l^\epsilon(x))$$

$$K_l^\epsilon(x) = E_n^\epsilon g'_\epsilon(x, \xi_l^\epsilon(x)) f_{xx}(x, t) g_\epsilon(x, \xi_l^\epsilon(x)) \\ - E g'_\epsilon(x, \bar{\xi}_l^\epsilon(x)) f_{xx}(x, t) g_\epsilon(x, \bar{\xi}_l^\epsilon(x))$$

$$L_l^\epsilon(x) = E_n^\epsilon f'_x(x, t) g_\epsilon(x, \xi_l^\epsilon(x))$$

$$M_{j,l}^\epsilon(x) = E_n^\epsilon (g'_\epsilon(x, \xi_j^\epsilon(x)) f_x(x, t))'_x g_\epsilon(x, \xi_l^\epsilon(x)) \\ - E (g'_\epsilon(x, \bar{\xi}_j^\epsilon(x)) f_x(x, t))'_x g_\epsilon(x, \bar{\xi}_l^\epsilon(x)).$$

The sums (9.4,5,6) below converge (absolutely) uniformly in ω, t, x, n and initial conditions for $P_n(x)$ in bounded sets, for each $f(\cdot) \in \mathcal{S}_0^{\infty, \infty}$.

$$(9.4) \quad \sum_{\ell=n}^{\infty} J_{\ell}^{\epsilon}(x), \quad \sum_{\ell=n}^{\infty} (J_{\ell}^{\epsilon}(x))_x, \text{ and also for } K, L \text{ replacing } J.$$

$$(9.5) \quad \sum_{\ell=n}^{\infty} (L_{\ell}^{\epsilon}(x))_{xx}$$

$$(9.6) \quad \sum_{\ell=n}^{\infty} \sum_{j=\ell+1}^{\infty} M_{j,\ell}^{\epsilon}(x), \quad \sum_{\ell=n}^{\infty} \sum_{j=\ell+1}^{\infty} (M_{j,\ell}^{\epsilon}(x))_x.$$

Remark. (A11)-(A12) hold if $\{\psi_n^{\epsilon}\}$ is mutually independent, \tilde{h}_{ϵ} and g are multiplicative in ξ and there are $|\alpha| < 1$ and twice continuously differentiable (in x) $b(\cdot, \cdot)$ such that

$$\xi_{\ell+1}^{\epsilon} = \alpha \xi_{\ell}^{\epsilon} + b(X_{\ell+1}^{\epsilon}, \psi_{\ell+1}^{\epsilon}), \quad E b(x, \psi_n^{\epsilon}) \equiv 0.$$

A Note on the Markov Case. For the case where $\{\psi_n^{\epsilon}\}$ are mutually independent for each $\epsilon > 0$, $\{X_n^{\epsilon}, \xi_{n-1}^{\epsilon}\}$ is a Markov process if Q_{ϵ} is merely a Borel function. Then E_n^{ϵ} in (A12) and in (9.7), (9.8) below can be replaced by conditioning on $X_n^{\epsilon}, \xi_{n-1}^{\epsilon}$ only. Instead of differentiating the functions in the sums in f_1^{ϵ} and f_2^{ϵ} with respect to x , as done in the proof of Theorem 3, it is only necessary that the $E_n^{\epsilon} f'_x(x, \epsilon n) \tilde{h}_{\epsilon}(x, \xi_{\ell}^{\epsilon}(x))$, $E g'_x(x, \xi_{\ell}^{\epsilon}(x)) f_{xx}(x, t) g_{\epsilon}(x, \xi_{\ell}^{\epsilon}(x))$, etc., be differentiated. Often the latter functions are smooth functions of x , uniformly for ξ_n^{ϵ} in bounded sets - even for Q_{ϵ} not satisfying the smoothness required by (A9). In that case, the proof of Theorem 3 can be carried out-but with derivatives of the conditional expectations

and expectations of the functions being taken in lieu of the derivatives of the functions themselves. Of course, the same can be said for the case of Theorem 3, if $\{\xi_n^\epsilon, \chi_n^\epsilon\}$ is Markov there.

To be more specific, In (9.2)-(9.6) replace (with the exception of the $I_{j,l}^\epsilon(x)$ and $M_{j,l}^\epsilon(x)$ terms) the $E(\cdot)_x$ and $E_n^\epsilon(\cdot)_x$ by $(E(\cdot))_l$ and $(E_n^\epsilon(\cdot))_x$, resp. Replace $M_{j,l}^\epsilon(x)$ and $I_{j,l}^\epsilon(x)$ by, resp.,

$$M_{j,l}^\epsilon(x) = E_n^\epsilon(E_l^\epsilon g'_\epsilon(x, \xi_j^\epsilon(x)) f_x(x, t))'_x g_\epsilon(x, \xi_l^\epsilon(x))]$$

$$- E[(E_l^\epsilon g'_\epsilon(x, \xi_j^\epsilon(x)) f_x(x, t))'_x g_\epsilon(x, \xi_l^\epsilon(x))],$$

$$I_{j,l}^\epsilon(x) = E[(E_l^\epsilon g'_\epsilon(x, \xi_j^\epsilon(x)) f_x(x, t))'_x g_\epsilon(x, \xi_l^\epsilon(x))]$$

$$- E[(E_l^\epsilon g'_\epsilon(x, \bar{\xi}_j^\epsilon(x)) f_x(x, t))'_x g_\epsilon(x, \bar{\xi}_l^\epsilon(x))],$$

where E_l^ϵ denotes conditioning on the Markov state at time l , whatever the process. Finally, replace the summand in (A7) by

$$E(E_0^\epsilon g_\epsilon(x, \bar{\xi}_l^\epsilon(x)))_x g_\epsilon(x, \bar{\xi}_0^\epsilon(x)).$$

Then, assuming that the above derivatives and conditional expectations and expectations are well defined, Theorem 4 continues to hold.

Theorem 4. Assume (A2)-(A12), where in (A4), (A6), (A7), $\{\bar{\xi}_n^\epsilon(x)$ replaces $\{\xi_n^\epsilon\}$ and the derivative in (A7) is a total

derivative⁺. Then the conclusion of Theorem 3 continues to hold.

Proof. The proof is the same as that of Theorem 3. The only change is that we use the following for the f_i^ϵ . First, $f_1^\epsilon(n\epsilon) = f_1^\epsilon(X_n^\epsilon, n\epsilon)$ where, analogously to the case in Theorem 3,

$$\begin{aligned} f_1^\epsilon(x, n\epsilon) = & \sqrt{\epsilon} \sum_{j=n}^{\infty} E_n^\epsilon f'_x(x, n\epsilon) g_\epsilon(x, \xi_j^\epsilon(x)) + \\ (9.7) \quad & + \epsilon \sum_{l=n}^{\infty} E_n^\epsilon f'_x(x, \epsilon n) \tilde{h}_\epsilon(x, \xi_l^\epsilon(x)) \\ & + \frac{\epsilon}{2} \sum_{l=n}^{\infty} [E_n^\epsilon g'_\epsilon(x, \xi_l^\epsilon(x)) f_{xx}(x, \epsilon n) g_\epsilon(x, \xi_l^\epsilon(x)) \\ & - E g'_\epsilon(x, \xi_l^\epsilon(x)) f_{xx}(x, \epsilon n) g_\epsilon(x, \xi_l^\epsilon(x))], \end{aligned}$$

and the initial condition for $\{\xi_l^\epsilon(X_n), l \geq n\}$, is $\xi_n^\epsilon(X_n) = \xi_n^\epsilon$.

Next, we use $f_2^\epsilon(n\epsilon) = f_2^\epsilon(X_n^\epsilon, n\epsilon)$, where

$$\begin{aligned} f_2^\epsilon(x, n\epsilon) = & \epsilon \sum_{l=n}^{\infty} \sum_{j=l+1}^{\infty} [E_n^\epsilon (f'_x(x, n\epsilon) g_\epsilon(x, \xi_j^\epsilon(x)))'_x g_\epsilon(x, \xi_l^\epsilon(x)) \\ (9.8) \quad & - E (f'_x(x, n\epsilon) g_\epsilon(x, \xi_j^\epsilon(x)))'_x g_\epsilon(x, \xi_l^\epsilon(x))] \end{aligned}$$

and with initial conditions

⁺I.e., $g_{\epsilon, x}(x, \xi_l^\epsilon(x))$ in (A7) is replaced by $g_{\epsilon, x}(x, \xi_l^\epsilon(x)) + g_{\epsilon, \xi}(x, \xi_l^\epsilon(x)) \xi_{l, x}^\epsilon(x)$.

$$\xi_n^\epsilon(X_n^\epsilon) = \xi_n^\epsilon, \quad \xi_{n+1,x}^\epsilon(X_n^\epsilon) = Q_x(X_n^\epsilon, \xi_n^\epsilon, \psi_{n+1}^\epsilon).$$

The forms of $\{\xi_n^\epsilon(x), \bar{\xi}_n^\epsilon(x)\}$ and $f_1^\epsilon, f_2^\epsilon$ are chosen so that, by following the method of Theorem 3, we get the correct cancellations and the centering about the correct - and intuitively reasonable - "mean-values" which make up the operator of the limit process.

10. An Example

A simple but interesting example will be given. It does not quite fit the format of Theorem 4, but can be handled by a very similar development, and it suggests some interesting possibilities, as well as indicating the potential power of the general method of the paper. We consider a very simple form of the Kiefer-Wolfowitz stochastic approximation for minimizing a function $q(x)$. Here $q(x) = \frac{1}{2} x' K x$, K positive definite and symmetric, and it is assumed that the derivative plus noise is observed, and the coefficient sequence is constant; i.e., the iterate sequence is

$$(10.1) \quad Y_{n+1}^\epsilon = Y_n^\epsilon - \epsilon [K Y_n^\epsilon + \psi_n],$$

where Y_n^ϵ is the n^{th} estimate of the minimum value 0 of $q(x)$ and the $\{\psi_n\}$ are mutually independent. It is occasionally suggested that, if observations at successive parameter points were averaged in some way, then convergence would be improved.

This question will now be investigated. A more general form for $q(\cdot)$ could be used, but we stick to the simple case in order to illustrate the main point as efficiently as possible.

Instead of (10.1), we deal with (10.2), where the successive observations are geometrically weighted for use in the iteration.

10. An Example

A simple but interesting example will be given. It does not quite fit the format of Theorem 4, but can be handled by a very similar development, and it suggests some interesting possibilities as well as indicating the potential power of the general method of the paper. We consider a very simple form of the Kiefer-Wolfowitz stochastic approximation for minimizing a function $d(x)$. Here $d(x) = \frac{1}{2} x' K x$, K positive definite and symmetric, and it is assumed that the derivative plus noise is observed, and the coefficient sequence is constant; i.e., the iterate sequence is

$$(10.1) \quad y_{n+1}^e = y_n^e - \epsilon [K y_n^e + v_n^e],$$

where y_n^e is the n th estimate of the minimum value 0 of $d(x)$ and the $\{v_n^e\}$ are mutually independent. It is occasionally suggested that, if observations at successive parameter points were averaged in some way, then convergence would be improved.

$$(10.2) \quad X_{n+1}^{\epsilon} = X_n^{\epsilon} - \epsilon \xi_n^{\epsilon}, \quad 0 \leq \alpha < 1, \quad \beta > 0,$$

$$\xi_n^{\epsilon} = \alpha \xi_{n-1}^{\epsilon} + \beta [KX_n^{\epsilon} + \psi_n], \quad \{\psi_n\} \text{ bounded, identically distributed.}$$

If $\alpha = 0$, then (10.2) reduces to (10.1). For $\alpha > 0$, there is some averaging. Define $X^{\epsilon}(\cdot)$ as in Section 1. The conditions of Theorem 4 all hold and by Theorem 4, $\{X^{\epsilon}(\cdot)\}$ is tight and converges (in $D^r[0, \infty)$, for some integer r) weakly to the solution of $\dot{x} = -(\beta K/(1-\alpha))x$.

Next, let us center $\{X_n^{\epsilon}\}$ about its "mean value" and examine the rate of convergence. Define $\hat{X}_n^{\epsilon}, \hat{\xi}_n^{\epsilon}, \tilde{X}_n^{\epsilon}, \tilde{\xi}_n^{\epsilon}$, by

$$\hat{X}_{n+1}^{\epsilon} = \hat{X}_n^{\epsilon} - \epsilon \hat{\xi}_n^{\epsilon}, \quad \hat{X}_0^{\epsilon} = X_0,$$

$$\hat{\xi}_n^{\epsilon} = \alpha \hat{\xi}_{n-1}^{\epsilon} + \beta K \hat{X}_n^{\epsilon}, \quad \hat{\xi}_0^{\epsilon} = \xi_0^{\epsilon},$$

$$\tilde{X}_{n+1}^{\epsilon} = \tilde{X}_n^{\epsilon} - \epsilon \tilde{\xi}_n^{\epsilon}, \quad \tilde{X}_0^{\epsilon} = 0,$$

$$\tilde{\xi}_n^{\epsilon} = \alpha \tilde{\xi}_{n-1}^{\epsilon} + \beta (K \tilde{X}_n^{\epsilon} + \psi_n), \quad \tilde{\xi}_0^{\epsilon} = 0,$$

then $X_n^{\epsilon} = \hat{X}_n^{\epsilon} + \tilde{X}_n^{\epsilon}$. Let $U_n^{\epsilon} = \tilde{X}_n^{\epsilon}/\sqrt{\epsilon}$. Then

$$(10.3) \quad U_{n+1}^{\epsilon} = U_n^{\epsilon} - \sqrt{\epsilon} \tilde{\xi}_n^{\epsilon}, \quad \tilde{U}_0^{\epsilon} = 0,$$

$$\tilde{\xi}_n^{\epsilon} = \alpha \tilde{\xi}_{n-1}^{\epsilon} + \beta (K \sqrt{\epsilon} U_n^{\epsilon} + \psi_n).$$

Now define $\{\tilde{\xi}_n^{\epsilon}(u)\}$ by

$$(10.4) \quad \tilde{\xi}_\ell^\epsilon(u) = \alpha \tilde{\xi}_{\ell-1}^\epsilon(u) + \beta(K\sqrt{\epsilon} u + \psi_\ell), \quad \ell \geq n, \text{ each } n,$$

where the initial conditions are assigned at n as they were for $\{\xi_\ell^\epsilon(x), \ell \geq n\}$ in Theorem 4. We analyze (10.3), which differs from the situation in Theorem 4 owing to the $\sqrt{\epsilon}$ in the $\tilde{\xi}_n^\epsilon$ equation and the fact that $^+ \bar{E} \tilde{\xi}_\ell^\epsilon(x)$ is not zero unless $u = 0$ (the bar $-$ denotes the associated stationary process). Nevertheless, an almost identical development to that of Theorems 3 or 4 can be carried out by following the technique of Theorem 3 and introducing $f_1^\epsilon, f_2^\epsilon$ with the correct centering, very similar to what was done in the discussion concerning Theorem 4. We simply state the result in the scalar case. $\{u^\epsilon(\cdot)\}$ is tight in $D[0, \infty)$ and converges weakly to the diffusion $U(\cdot)$:

$$(10.5) \quad \begin{aligned} dU &= \frac{-\beta K}{(1-\alpha)} U dt + C dW, \quad U(0) = 0, \\ C^2 &= \frac{\sigma^2 \beta^2}{(1-\alpha^2)} \left[1 + \frac{2\alpha}{1-\alpha} \right] \end{aligned}$$

where $\sigma^2 = E\psi_n^2$. If $\{U_n^\epsilon, \epsilon > 0, n > 0\}$ is tight, then so is $\{U^\epsilon(t_\epsilon + \cdot), \epsilon > 0\}$ for any sequence $t_\epsilon \rightarrow \infty$, and the weak limit of the last sequence is the stationary solution to (10.5).

The stationary variance of (10.5) is $C^2(1-\alpha)/2\beta K$ or

$$\sigma^2 \beta / 2K(1-\alpha).$$

If $\beta = (1-\alpha)$ then the mean rate of convergence as well as the asymptotic normalized variance do not depend on α . For smaller β , the mean rate is lower, the rate of decrease of the covariance of $U(\cdot)$ slower, but the

⁺ Here, $g_\epsilon(u, \xi) = \xi$.

asymptotic normalized variance is smaller. So there is no clear advantage to averaging, although other schemes may yield different results. The format developed in this paper seems to be quite suitable for the analysis of such problems.

- [1] G. A. G. (1972). Diffusion approximation for stochastic processes. *Ann. Prob.* 1, 58-74.
- [2] D. W. Varadhan, G. A. G. (1972). Multidimensional diffusion approximation for stochastic processes. *Ann. Prob.* 1, 58-74.
- [3] T. G. Kurtz (1972). Limit theorems of Markov processes. *Ann. Prob.* 1, 58-74.
- [4] Papanicolaou, G. A., Kushner, H. J. (1974). Asymptotic theory of mixing stochastic ordinary differential equations. *Comm. Pure and Appl. Math.* 27, 641-688.
- [5] Kushner, H. J. (1978). Jump-diffusion approximations for ordinary differential equations with random right hand sides, to appear *SIAM J. on Control and Optimization*, Autumn, 1978.
- [6] Kushner, H. J. (1979). A martingale method for the convergence of a sequence of processes to a jump diffusion process, submitted to *Z. Wahrscheinlichkeitstheorie*.
- [7] Billingsley, P. (1968). *Convergence of Probability Measures*, John Wiley and Sons, New York.
- [8] Kushner, H. J. (1979). Diffusion approximations to output processes of non-linear systems with wide band inputs, and applications, submitted to *IEEE Trans. on Information Theory*.
- [9] Blankenship, G. A., Papanicolaou, G. A. (1978). Stability and control of stochastic systems with wide band noise disturbances, *SIAM J. on Appl. Math.* 34, pp. 411-470.
- [10] Kushner, H. J., Han Huang (1978). Rates of convergence for stochastic approximation type algorithms, *SIAM J. on Control and Optimization*, September, 1978.
- [11] Kushner, H. J., Han Huang (1978). Asymptotic properties of stochastic approximations with constant coefficients (rates of convergence), submitted to *SIAM J. on Control and Optimization*.
- [12] Papanicolaou, G. A., Strook, D. W., Varadhan, S. R. S., A. Martinelli (1978). Approach to some limit theorems, *Proceedings, Duke University Conference on Turbulence*, 1978.

REFERENCES

- [1] Guess, H.A., Gillespie, J.H. (1977). Diffusion approximations to linear stochastic difference equations with stationary coefficients, *J. Appl. Prob.* 14, 58-74.
- [2] Strook, D.W., Varadhan, S.R.S. (1979). Multidimensional Diffusion Processes, Springer, Berlin.
- [3] Kurtz, T.G. (1975). Semigroups of conditioned shifts and approximations of Markov processes, *Ann. Prob.* 3, 618-642.
- [4] Papanicolaou, G.C., Kohler, W. (1974). Asymptotic theory of mixing stochastic ordinary differential equations, *Comm. Pure and Appl. Math.* 27, 641-668.
- [5] Kushner, H.J. (1978). Jump-diffusion approximations for ordinary differential equations with random right hand sides, to appear *SIAM J. on Control and Optimization*, Autumn, 1979.
- [6] Kushner, H.J. (1979). A martingale method for the convergence of a sequence of processes to a jump diffusion process, submitted to *Z. Wahrscheinlichkeitstheorie*.
- [7] Billingsley, P. (1968). Convergence of Probability Measures, John Wiley and Sons, New York.
- [8] Kushner, H.J. (1979). Diffusion approximations to output processes of non-linear systems with wide band inputs, and applications, submitted to *IEEE Trans. on Information Theory*.
- [9] Blankenship, G., Papanicolaou, G.C. (1978). Stability and control of stochastic systems with wide band noise disturbances, *SIAM J. on Appl. Math.*, 34, pp. 437-476.
- [10] Kushner, H.J., Hai Huang (1978). Rates of convergence for stochastic approximation type algorithms, *SIAM J. on Control and Optimization*, September, 1979.
- [11] Kushner, H.J., Hai Huang (1979). Asymptotic properties of stochastic approximations with constant coefficients (Rates of convergence), submitted to *SIAM J. on Control and Optimization*.
- [12] Papanicolaou, G.C., Strook, G., Varadhan, S.R.S., "A martingale approach to some limit theorems", *Proceedings, Duke University Conference on Turbulence*, 1976.